Nombre:	CI:	Carrera

EXAMEN

4 de Marzo de 2015

Ejercicio 1 (20 puntos)

Suponga que la probabilidad de que un individuo sea afectado por un cierto virus es 0.02. Se sabe además que entre los individuos afectados por el virus la probabilidad de que se laven las manos regularmente es 0.3, mientras que en entre los individuos no afectados esta probabilidad es de 0.6.

1. Calcular la probabilidad de que un individuo cualquiera se lave regularmente las manos.
2. ¿Cuál es la probabilidad de que un individuo se lave regularmente las manos y esté afectado por el virus?
3. Asumiendo que un individuo se lava las manos regularmente, ¿Cuál es la probabilidad de que esté afectado por el virus ?
4. Calcular la probabilidad de que un individuo esté afectado siendo que no se lava las manos regularmente.

Solución Ejercicio 1

Se definen los eventos

$$V = \{\text{estar afectado por el virus}\}$$

$$L = \{\text{lavarse las manos regularmente}\}.$$

Se sabe que: $P(V) = 0.02$, $P(V^c) = 1 - P(V) = 0.98$, $P(L\,|V) = 0.3$ y $P(L\,|V^c) = 0.6$, donde V^c es el evento complemento de V.

1. La probabilidad de que un individuo cualquiera se lave regularmente las manos se puede calcular usando la fórmula de la probabilidad total:

$$P(L) = P(L\,|V) \times P(V) + P(L\,|V^c)\,P(V^c) = 0.3 \times 0.02 + 0.6 \times 0.98 = 0.594.$$

2. La probabilidad de que un individuo se lave regularmente las manos y esté afectado por el virus es por definición:

$$P(L \cap V) = P(L\,|V)\,P(V) = 0.3 \times 0.02 = 0.006.$$

3. Usando la fórmula de Bayes, resulta que la probabilidad de que un individuo esté afectado por el virus dado que se lava las manos regularmente es:

$$P(V\,|L) = \frac{P(L\,|V)\,P(V)}{P(L)} = \frac{0.3 \times 0.02}{0.594} = 0.01.$$

4. Análogamente la probabilidad de que un individuo esté afectado por el virus siendo que no se lava las manos regularmente es:

$$P(V\,|L^c) = \frac{P(V \cap L^c)}{P(L^c)} = \frac{0.014}{1 - 0.594} = 0.034$$

donde $P(V \cap L^c) = P(V) - P(V \cap L)$.

Ejercicio 2 (40 puntos)

1. La bacteria *Helicobacter Pylori* genera deficiencias en la absorción de nutrientes debido a su interferencia en la secreción de ácidos por parte del estómago. Sea X la variable aleatoria que representa el tiempo de vida de dicha bacteria (medido en días). Se asume que la distribución de X es exponencial de parámetro $\lambda = 0.1$.

   (a) Indicar cuáles son los valores posibles que puede tomar la variable aleatoria X.

   (b) i. Calcular la probabilidad de que el tiempo de vida de la bacteria sea mayor a una semana.

   ii. Si se sabe que la bacteria ya vivió dos días, ¿cuál es la probabilidad de que viva más de una semana?

   (c) Indicar $\mathrm{E}(X)$ y $\mathrm{Var}(X)$.

2. Para tratar la infección producida por dicha bacteria existen dos tratamientos posibles. El primero es un tratamiento tradicional a base de antibióticos, mientras el segundo consiste en la aplicación de una técnica sofisticada denominada *cuádruple terapia*. En ambos casos se asume que la distribución luego del tratamiento sigue siendo exponencial con parámetros λ_1 y λ_2 respectivamente.

   (a) Para investigar la efectividad del primer tratamiento (si se reduce significativamente el tiempo de vida medio de la bacteria), este tratamiento se aplicó en 200 pacientes infectados. Se obtuvieron los siguientes datos del ciclo de vida de la bacteria:

   $$\sum_{i=1}^{200} x_i = 1182.25 \qquad \text{y} \qquad \sum_{i=1}^{200} x_i^2 = 13239.14,$$

   donde x_i es el tiempo de vida medio de la bacteria en el i-ésimo paciente.

   i. Realizar un test de hipótesis que permita concluir sobre la eficacia del tratamiento.

   ii. Dar un intervalo de confianza al 95% para λ_1.

   iii. Suponiendo como verdadera la estimación de la varianza obtenida a partir de la muestra anterior, hallar el número mínimo de pacientes que deberían ser tratados para obtener un error en la estimación del tiempo medio de vida de la bacteria menor a 0.5 con probabilidad mayor a 0.95.

   (b) Para el segundo tratamiento no se tienen datos tan precisos, pero se sabe que se aplicó a un gran número de personas y que se obtuvo que en un 50% de los casos la bacteria tuvo un tiempo de vida menor a 4 días. Estimar el parámetro λ_2.

Solución Ejercicio 2

1. (a) Como la variable aleatoria X se distribuye exponencial de parámetro $\lambda = 0.1$ se sabe que los valores que puede obtener es $[0, +\infty)$.

   (b) i. $P(X > 7) = 1 - P(X \leq 7) = e^{-\lambda_1 7} = e^{-0.7} = 0.497$.

   ii. $P(X > 7 | X > 2) = P(X > 5) = e^{-\lambda_1 5} = e^{-0.5} = 0.606$.

   (c) Se sabe que

$$\mathrm{E}(X) = \frac{1}{\lambda} = 10$$
$$\mathrm{Var}(X) = \frac{1}{\lambda^2} = 100$$

2. (a) i. Se desea saber si el tratamiento reduce el tiempo medio de vida de la bacteria. Por lo tanto el test de hipótesis a realizar en la media es:

$$\begin{cases} H_0 : \mu_1 \geq \mu = \frac{1}{\lambda} = 10 \\ H_1 : \mu_1 < \mu \end{cases}$$

que es equivalente a plantear el siguiente test sobre el párametro λ:

$$\begin{cases} H_0 : \lambda_1 \leq \lambda \\ H_1 : \lambda_1 > \lambda \end{cases}$$

La región crítica para este test es $\mathcal{R}_\alpha = \left\{ \sqrt{n} \left(\frac{\bar{X}_n - 10}{s_n} \right) \leq z_\alpha \right\}$. De los datos se obtiene que

$$\hat{\mu} = \frac{1}{200} \sum_{i=1}^{200} x_i = 5.91$$

$$S_n^2 = \frac{1}{199} \sum_{i=1}^{200} x_i^2 - \frac{200}{199} 5.91^2 = 31.41.$$

Luego $E = \sqrt{n} \left(\frac{\bar{X}_n - 10}{s_n} \right) = -10.32$ y el p-valor es $\alpha^* = \Phi(E) = 0$. Por lo tanto se rechaza la hipótesis nula.

ii. Planteando un intervalo de confianza aproximado para la media se tiene que:

$$I_\mu = \left[\hat{\mu} - \frac{z_{0.975} s_n}{\sqrt{200}}, \hat{\mu} + \frac{z_{0.975} s_n}{\sqrt{200}} \right] = [5.91 - 0.78, 5.91 + 0.78] = [5.13, 6.69].$$

Usando que la esperanza de una exponencial es $1/\lambda$ obtenemos el intervalo de confianza para λ:

$$I_\lambda = \left[\frac{1}{6.69}, \frac{1}{5.13} \right] = [0.15, 0.19].$$

iii. Para que el error del tiempo medio de vida de la bacteria sea menor a 0.5 con probabilidad mayor a 95%, necesitamos hallar n tal que la diferencia entre la estimacin y la media sea menor que 0.5. Esto es, el largo total del intervalo de confianza tiene que ser igual a 1. Asumiendo que s_n coincide con el hallado en la muestra anterior, necesitamos que

$$\frac{z_{0.975}s_n}{\sqrt{n}} \leq 0.5,$$

de donde se tiene que

$$\sqrt{n} \geq \frac{z_{0.975}s_n}{0.5} = 22 \Rightarrow n \geq 484$$

El mínimo $n = 484$.

(b) De la información brindada por la muestra se tiene que $P(X < 4) \sim 0.5$. Dado que sigue una distribución exponencial, se tien que $P(X < 4) = 1 - e^{-\lambda_2 4}$. De igualar ambar ecuaciones resulta que un estimador de λ_2 es $\hat{\lambda}_2 = 0.17$.

Ejercicio 3 (38 puntos)

Se desea comparar la actividad cerebral de jóvenes y adultos mayores mientras manejan. Para hacerlo se coloca a los individuos de una muestra test en un simulador de manejo y se les realiza un electroencefalograma al mismo tiempo. Para cada individuo se registra la frecuencia de las ondas (medidas en Hz) durante la experiencia de manejo. Los datos obtenidos son los siguientes:

Adultos Mayores	11.5	11.9	8.6	9.5	7.5	8.0	6.1	6.5	8.7	6.6
Jóvenes	5.5	5.1	5.9	4.4	4.7	4.3	5.4	3.9	4.5	6.4

Se asume que cada una de las muestras anteriores es una muestra aleatoria simple (iid).

1. ¿Se puede afirmar que las muestras son independientes? Realizar un test de hipótesis.
2. ¿Es posible afirmar que la actividad cerebral en jóvenes y adultos mayores durante el manejo es la misma? Realizar un test de hipótesis.
3. Asumiendo que los datos son gaussianos ¿Es posible afirmar que la actividad cerebral en un grupo es mayor a la del otro grupo?
4. Ahora se decide medir la capacidad de reacción en ambos grupos. Para ello en el simulador se agrega un obstáculo y se registra si el individuo fue capaz de esquivarlo o no. Se observó que dentro del grupo de los individuos jóvenes 20 individuos lograron esquivar el obstáculo mientras que 25 no lo hicieron. En el grupo de adultos mayores, 10 individuos lograron esquivarlo y 45 no. ¿Es posible afirmar que la edad influye en la capacidad de reacción? Realizar un test de hipótesis.

Para el ejercicio 3, se asume como p-valor $\alpha^* = 0.1$.

Solución Ejercicio 3

1. Para estudiar independencia entre dos muestras se plantea el test de correlación de rangos de Spearman (para dos muestras). El estadístico de prueba es:

$$R_S = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^{n} (R(A_i) - R(J_i))^2$$

Entonces:

A_i	J_i	$R(A_i)$	$R(J_i)$	$(R(A_i) - R(J_i))^2$
11,5	5,5	9	8	1
11,9	5,1	10	6	16
8,6	5,9	6	9	9
9,5	4,4	8	3	25
7,5	4,7	4	5	1
8,0	4,3	5	2	9
6,1	5,4	1	7	36
6,5	3,9	2	1	1
8,7	4,5	7	4	9
6,6	6,4	3	10	49

Por lo tanto se tiene que:

$$R_S = 1 - \frac{6}{10 \times 99}(1 + 16 + 9 + 25 + 1 + 9 + 36 + 1 + 9 + 49) = 0.055$$

Como $R_S > 0$ planteamos:

$$\begin{cases} H_0: & \text{A y J son independientes} \\ H_1: & \text{Hay correlación positiva} \end{cases}$$

De la tabla del test se tiene que $\alpha^* = 0,446 > 0,1$ por lo tanto no se rechaza H_0.

2. Para comparar las distribuciones de los dos grupos, planteamos el test de Kolmogorov-Smirnov para dos muestras, es decir:

$$\begin{cases} H_0: & A \sim J \\ H_1: & \text{no } H_0 \end{cases}$$

El estadístico es:

$$D_{mn} = \sup_{t \in \mathbb{R}} |F_m^A(t) - F_n^J(t)|$$

Muestra	Muestra conjunta ordenada	F_m^A	F_n^J	$\lvert F_m^A - F_n^J \rvert$
J	3,9	0	0,1	0,1
J	4,3	0	0,2	0,2
J	4,4	0	0,3	0,3
J	4,5	0	0,4	0,4
J	4,7	0	0,5	0,5
J	5,1	0	0,6	0,6
J	5,4	0	0,7	0,7
J	5,5	0	0,8	0,8
J	5,9	0	0,9	0,9
A	6,1	0,1	0,9	0,8
J	6,4	0,1	1	0,9
A	6,5	0,2	1	0,8
A	6,6	0,3	1	0,7
A	7,5	0,4	1	0,6
A	8,0	0,5	1	0,5
A	8,6	0,6	1	0,4
A	8,7	0,7	1	0,5
A	9,5	0,8	1	0,3
A	11,5	0,9	1	0,1
A	11,9	1	1	0

Por lo tanto $D_{mn} = 0.9$ y $nmD_{mn} = 90$. Usando la tabla de K-S para dos muestras se tiene que $\alpha^* < 0,010$ por lo tanto se rechaza la hipótesis nula.

3. Hay al menos dos maneras de resolver este ejercicio: usando un test no párametrico de corrimientos (por ej. MWW) o usando intervalos de confianza.

Planteando Mann Whitney Wilcoxon (MWW):

$$\begin{cases} H_0: & A \sim J \\ H_1: & A \sim J + \theta, \theta \in \mathbb{R} \end{cases}$$

Muestra	Muestra conjunta ordenada	Rango
J	3,9	1
J	4,3	2
J	4,4	3
J	4,5	4
J	4,7	5
J	5,1	6
J	5,4	7
J	5,5	8
J	5,9	9
A	6,1	10
J	6,4	11
A	6,5	12
A	6,6	13
A	7,5	14
A	8,0	15
A	8,6	16
A	8,7	17
A	9,5	18
A	11,5	19
A	11,9	20

El estadístico del test es la suma de los rangos de la muestra de menor tamaño (en este caso da lo mismo) en la muestra conjunta ordenada.

$$T_A = 10 + 12 + 13 + 14 + 15 + 16 + 17 + 18 + 19 + 20 = 154$$

Resulta que $T_A \geq \frac{10\times(10+10+1)}{2}$, por lo cuál se plantea el test a una cola:

$$\begin{cases} H_0 : & A \sim J \\ H_1 : & A \sim J + \theta, \theta > 0 \end{cases}$$

Usando la tabla de MWW se obtiene que el p-valor es $\alpha^* = 0$, por lo tanto se rechaza H_0 y podemos decir que la actividad cerebral de los adultos mayores es mayor que la de los jóvenes.

Planteando intervalos de confianza: Como las muestras son gaussianas y σ es desconocido, los intervalos de confianza son de la forma

$$\left[\bar{X_n} - \frac{s_n}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1), \bar{X_n} + \frac{s_n}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1)\right]$$

donde $t(n-1)$ es la distribución t- student con $n-1$ grados de libertad.

Tomando un intervalo de confianza al 95%, se tiene que $t_{\frac{\alpha}{2}}(n-1) = t_{0.025}(9) = 2.262$

Para la muestra de adultos mayores:

$\bar{A_n} = \frac{1}{10} \sum\limits_{i=1}^{10} A_i = 8.49$ y $s_n^2 = \frac{1}{n-1} \sum\limits_{i=1}^{10} (A_i - \bar{A_n})^2 = 4.$

Por lo tanto el intervalo de confianza es: $\left[8.49 - \frac{2}{\sqrt{10}}2.262, 8.49 + \frac{2}{\sqrt{10}}2.262\right] = [7.1, 9.9]$.

Para la muestra de jóvenes:

$\bar{J}_n = 5.01$ y $s_n = 0.79$, de donde el intervalo de confianza es $[4.4, 5.8]$.

Como los intervalos de confianza son disjuntos, podemos decir que la media de la actividad cerebral de los adultos mayores es mayor que la media de la actividad cerebral de los jvenes.

4. Investigar si la edad influye o no en la capacidad de reacción de los individuos, equivale a investigar si los tiempos de reacción son independientes de la edad. Para esto, y considerando los datos que se tienen se aplicará un test Chi- cuadrado (tabla de contingencia).

$$\begin{cases} H_0: & X \text{ e } Y \text{ son independientes} \\ H_1: & \text{no } H_0 \end{cases}$$

Tenemos entonces la siguiente tabla de contingencia:

	Logra esquivarlo	No logra esquivarlo	Total
Adulto Mayor	$N_{11} = 10$	$N_{12} = 45$	$N_{1*} = 55$
Joven	$N_{21} = 20$	$N_{22} = 25$	$N_{2*} = 45$
Total	$N_{*1} = 30$	$N_{*2} = 70$	$n = 100$

El estadístico de prueba es:

$$E = \frac{n\left(|N_{11}N_{22} - N_{12}N_{21}| - \frac{n}{2}\right)^2}{N_{1*} \times N_{*1} \times N_{*2} \times N_{2*}} = \frac{45\left(|10 \times 25 - 45 \times 20| - \frac{100}{2}\right)^2}{55 \times 30 \times 70 \times 45} = 3.67$$

La región crítica es:

$$\mathcal{R}_\alpha = \left\{E > \chi^2_\alpha((r-1)(s-1))\right\}$$

En nuestro caso $r = s = 1$ entonces usando la tabla de χ^2 se tiene que $0,05 < \alpha* < 0,1$. Por lo tanto se rechaza H_0. Es decir que podemos suponer que hay relación entre la edad y la capacidad de reacción.